

ZMATH 2016f.01257

Kim, Albert Y.; Escobedo-Land, Adriana

OkCupid data for introductory statistics and data science courses.

J. Stat. Educ. 23, No. 2, 25 p., electronic only (2015).

Summary: We present a data set consisting of user profile data for 59,946 San Francisco OkCupid users (a free online dating website) from June 2012. The data set includes typical user information, lifestyle variables, and text responses to 10 essay questions. We present four example analyses suitable for use in undergraduate introductory probability and statistics and data science courses that use R. The statistical and data science concepts covered include basic data visualization, exploratory data analysis, multivariate relationships, text analysis, and logistic regression for prediction.

Classification: K45 K85

Keywords: stochastics; statistics; university teaching; introductory statistics courses; data sets; online dating; data science; big data; logistic regression; text mining; exploratory data analysis; visualization; multivariate relationships

<http://ww2.amstat.org/publications/jse/v23n2/kim.pdf>